

Mata Kuliah : IBDA3322 / Natural Language Processing
Hari, Tanggal : XX, XX
Waktu : XX
Sifat Ujian : Proyek Kelompok (1 kelompok maksimal 3 orang)

PETUNJUK

- Kerjakanlah proyek di bawah ini lalu dikumpulkan dalam format .ipynb dengan mengikuti rubrik yang ada lalu dikumpulkan di dalam Canvas bagian *assignments*.
- Setelah selesai mengerjakan ujian, tuliskan komponen persentase kontribusi dari setiap anggota kelompok, selanjutnya salinlah pernyataan komitmen integritas berikut ini pada file .ipynb:

"Di hadapan TUHAN yang hidup, saya menegaskan bahwa saya tidak memberikan maupun menerima bantuan apapun—baik lisan, tulisan, maupun elektronik—di dalam ujian ini selain daripada apa yang telah diizinkan oleh pengajar, dan tidak akan menyebarkan baik soal maupun jawaban ujian kepada pihak lain."

PENGANTAR

Terdapat banyak sekali masalah di sekitar kita yang dapat kita selesaikan dengan NLP. Oleh karena itu peserta kelas NLP diminta untuk berpartisipasi aktif dengan menggunakan ilmu yang sudah diperoleh di kelas ini untuk menawarkan solusi pemodelan bagi permasalahan tertentu. Kalian bebas menggunakan konsep apapun dari topik yang sudah dipelajari di kelas. Solusi yang ditawarkan perlu dibuat dalam bentuk laporan pemrograman pada file notebook (ipynb) dan dipresentasikan.

DESKRIPSI TUGAS

1. Proposal (20 poin)

Sebelum melakukan pemodelan, diperlukan pertimbangan yang matang tentang data apa yang dipilih dan pemodelan seperti apa yang diperlukan. Maka 1 minggu sebelum proyek ini dimulai, kumpulkanlah melalui text entry yang sudah disiapkan di canvas, sebuah proposal singkat (kurang dari 500 kata) mengenai masalah yang ingin diselesaikan beserta data dan pemodelan yang akan kalian gunakan. Proposal akan dinilai secara kualitatif terlebih dahulu untuk memastikan proyek dari setiap kelompok unik. Jika terdapat pilihan data yang sama, hanya proposal dari kelompok yang mengumpulkan terlebih dahulu yang diterima. Kalian diperbolehkan menggunakan algoritma yang dipakai dalam tugas-tugas individu, namun **tidak boleh mengerjakan proyek yang sama atau menggunakan data yang sama**. Kalian bebas menggunakan library python apapun (tensorflow/pytorch/trax/keras). Jika proposal lengkap dan valid, maka otomatis akan mendapatkan nilai 20 poin, dimana pembagiannya adalah sebagai berikut:

- (5 poin) masalah yang coba diselesaikan
- (5 poin) sumber data (contoh: url_link/survey/SEP/BEM) dan deskripsi singkat data
- (5 poin) cara memperoleh data (contoh: crawling/scraping/raw data ready)
- (5 poin) jenis model NLP yang akan digunakan (contoh: MLP/LSTM/Attention/transformer/BERT/GPT/dan lain-lain)

2. Notebook dan Presentasi (80 poin + 10 poin bonus)

Dengan menggunakan Jupyter Notebook lakukanlah proses pengolahan data dan pemodelan untuk masalah yang kelompok kalian pilih. Kalian juga diminta menjelaskan secara lebih detail pekerjaan yang dilakukan dengan memanfaatkan fitur markdown dari Jupyter Notebook supaya diperoleh tampilan yang lebih mudah dibaca. Kelompok kalian akan mempresentasikan proyek ini dalam bentuk jupyter notebook di depan kelas pada minggu UAS.

Presentasi kalian akan dinilai berdasarkan:

- Alur penjelasan mengenai permasalahan yang dihadapi beserta kesesuaiannya dengan data yang dipilih dan algoritma yang digunakan (10 poin).
- Kompleksitas arsitektur yang dipilih:
 - o Tidak ada arsitektur (0 poin)
 - o Menggunakan prosedur non-machine learning (hanya 5 poin)
 - o Menggunakan machine learning klasik (hanya 10 poin)
 - o Menggunakan machine learning dengan tambahan embedding (hanya 15 poin)
 - o Menggunakan deep learning sederhana seperti MLP (hanya 20 poin)
 - o Menggunakan sequence models (hanya 25 poin)
 - o Menggunakan mekanisme attention (hanya 30 poin)
 - o Menggunakan vanilla transformer (hanya 35 poin)
 - o Menggunakan modifikasi dari transformer seperti BERT/GPT (full 40 poin)
- Keberhasilan algoritma untuk dijalankan baik dalam proses training maupun testing (10 poin).
- Tanya jawab (20 poin).
- Nilai bonus: mengunggah hasil proyek ke github (10 poin).